The consequences of demographic stochasticity on fixation

Diala Abu Awad, Camille Coron

CBGP - 21 Feb. 2017

The fate of mutant genes

Evolutionary forces shape the genetic diversity of populations

- Mutation
- Selection
- Drift
- Migration

The fate of mutant genes

Evolutionary forces shape the genetic diversity of populations

- Mutation
- **Selection**
- **Drift**
- Migration

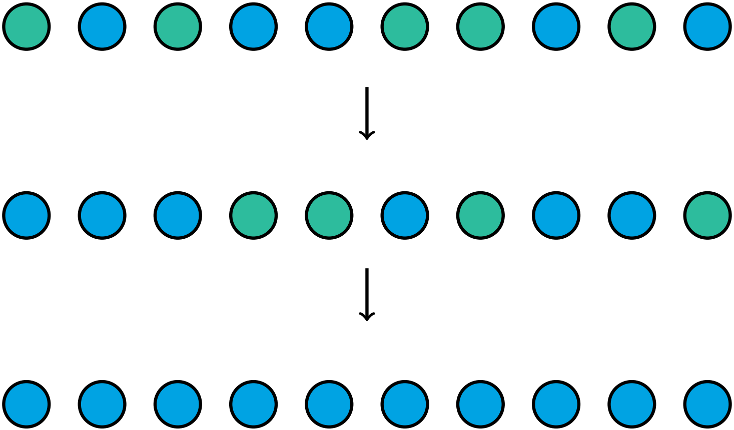
The fate of mutant alleles

Evolutionary forces shape the genetic diversity of populations

- Adaptation (fixation of benefical alleles)
- Selection against deleterious alleles
- Neutral diversity (adaptive potential)

The Wright-Fisher model (Genetic drift)

Two absorbing states:



The Wright-Fisher Diffusion

Kimura's diffusion of the Wright-Fisher model

- Probability of fixation
- Time to fixation

The Wright-Fisher Diffusion

A robust method despite strong underlying assumptions:

- Fixed population size
- Panmixia
- Non-overlapping generations

The Wright-Fisher Diffusion

It has been generalised to account for a variety of complications:

- Deterministically varying size (independently of genotypes) i.e. Otto and Whitlock 1997
- Inbreeding i.e. Caballero and Hill 1992
- Structured populations i.e. Roze and Rousset 2004
- etc...

Varying population size: A stochastic process

In natural populations:

- Population size varies stochastically
  - Demographic stochasticity
  - Environmental instability

Varying population size: A stochastic process

In natural populations:

- Population size varies stochastically
  - Demographic stochasticity
  - Environmental instability
- There is a potential feed-back between genotypes and population size (i.e. selection for more competitive and/or more fertile individuals)

Varying population size: A stochastic process

It has been shown that the harmonic mean of population size suffices in models with deterministically (and neutrally) varying size (Ewens 1967, Kimura 1970, Otto and Whitlock 1997)

$\rightarrow$ Is this still valid in stochastically varying populations (with feed-back)?

General model

- Diploid individuals
- Single bi-allelic locus : *AA Aa aa*
- Population size is a variable and not a parameter

Rescaled birth-and-death process

At each time t the population is represented by a vector (with 1,2,3 representing *AA* *Aa* and *aa* respectively)

$$\left(\mathbf{Z}_t^K\right)_{t\geq 0} = \left(Z_t^{1,K}, Z_t^{2,K}, Z_t^{3,K}\right)_{t\geq 0}$$

which gives the respective number of individuals of each type, divided by K (a scaling parameter that goes to infinity). (Fournier and Meleard 2004; Champagnat and Meleard 2007; Collet, Meleard and Metz 2012; Coron 2014

Rescaled birth-and-death process

If the population is at a state $\mathbf{z} = (z_1, z_2, z_3)$, the birth rates $\lambda_i^K(\mathbf{z})$ for all $i \in \{1, 2, 3\}$ model sexual Mendelian reproduction

$$\lambda_1^K(\mathbf{z}) = Kb_1^K \left[\alpha\Big(z_1 + \frac{z_2}{4}\Big) + (1-\alpha)\frac{(z_1 + \frac{z_2}{2})^2}{n} \right],$$

$$\lambda_2^K(\mathbf{z}) = Kb_2^K \left[\alpha\frac{z_2}{2} + (1-\alpha)2\frac{(z_1 + \frac{z_2}{2})(z_3 + \frac{z_2}{2})}{n} \right],$$

$$\lambda_3^K(\mathbf{z}) = Kb_3^K \left[\alpha\Big(z_3 + \frac{z_2}{4}\Big) + (1-\alpha)\frac{(z_3 + \frac{z_2}{2})^2}{n} \right].$$

with $n = z_1 + z_2 + z_3 \neq 0$

Rescaled birth-and-death process

If the population is at a state z, the rate $\mu_i^K(z)$ at which an individual with genotype i dies in the population is then given by:

$$\mu_1^K(z) = Kz_1(d^K + K(c^K z_1 + c^K z_2 + c^K z_3)),$$
$$\mu_2^K(z) = Kz_2(d^K + K(c^K z_1 + c^K z_2 + c^K z_3)),$$
$$\mu_3^K(z) = Kz_3(d^K + K(c^K z_1 + c^K z_2 + c^K z_3)).$$

The demographic parameter d^K (resp. $c^K > 0$) is the intrinsic death rate (resp. the competition rate) of individuals.

Rescaled birth-and-death process

The demographic parameters b^K, d^K and c^K are scaled both by K and a parameter γ, the latter scaling the speed with which births and deaths occur, giving:

$$b_1^K = \gamma K + \rho,$$
$$b_2^K = \gamma K + \rho + h\sigma,$$
$$b_3^K = \gamma K + \rho + \sigma,$$

and

$$d^K = \gamma K \quad \text{and} \quad c^K = \frac{\xi}{K}.$$

Rescaled birth-and-death process

- We follow the evolution of population mass not size with : $\mathcal{N}_t^K = Z_t^{1,K} + Z_t^{2,K} + Z_t^{3,K}$
- For large K the effect of the selection coefficient σ on b_i^K is inherently weak, but it will still have a macroscopic effect on population mass.

Limiting the diffusion process

The limiting population dynamics can be represented at time t by the couple $(\mathcal{N}_t^K, X_t^K)$ giving the population size and the proportion of allele a. Coron 2014

$$d\mathcal{N}_t = \sqrt{2\gamma\mathcal{N}_t}dB_t^1 \tag{1a}$$

$$+\mathcal{N}_t\Big[\rho - \xi\mathcal{N}_t + \sigma X_t\Big(2h + X_t(1-2h) + F(1-X_t)(1-2h)\Big)\Big]dt,$$

$$dX_t = \sqrt{\frac{2\gamma X_t(1-X_t)}{2\frac{\mathcal{N}_t}{1+F}}}dB_t^2 \tag{1b}$$

$$+\sigma X_t(1-X_t)\Big[h + X_t(1-2h) + F(1-X_t-h+2X_t h)\Big]dt.$$

where $(B_t^1, B_t^2)_{t\geq 0}$ is a bi-dimensional standard Brownian motion.

Limiting the diffusion process

The limiting population dynamics can be represented at time t by the couple $(\mathcal{N}_t^K, X_t^K)$ giving the population size and the proportion of allele a. Coron 2014

$$d\mathcal{N}_t = \sqrt{2\gamma\mathcal{N}_t}dB_t^1$$

$$+\mathcal{N}_t\Big[\rho - \xi\mathcal{N}_t + \sigma X_t\Big(2h + X_t(1-2h) + F(1-X_t)(1-2h)\Big)\Big]dt,$$

$$dX_t = \sqrt{\frac{2\gamma X_t(1-X_t)}{2\frac{\mathcal{N}_t}{1+F}}}dB_t^2$$

$$+\sigma X_t(1-X_t)\Big[h + X_t(1-2h) + F(1-X_t - h + 2X_t h)\Big]dt.$$

where $(B_t^1, B_t^2)_{t\geq 0}$ is a bi-dimensional standard Brownian motion.

Limiting the diffusion process

By setting γ to $1/2$:

$$dX_t = \sqrt{\frac{X_t(1-X_t)}{2\frac{\mathcal{N}_t}{1+F}}}dB_t^2$$
$$+\sigma X_t(1-X_t)\Big[h + X_t(1-2h) + F(1-X_t-h+2X_t h)\Big]dt.$$

We have the same expression for changes in allelic frequencies as in Caballero and Hill (1992)

Simulations run

- Analytical approximations could not be made (bi-dimensional process)
- Numerical results were obtained using simulations of equations (5a) and (5b) were simulated using a script written in C++
- Simulations for fixed population size were also run

Demographic stochasticity

Neutral case ($\sigma = 0$): Population mass is independent of its genetic composition

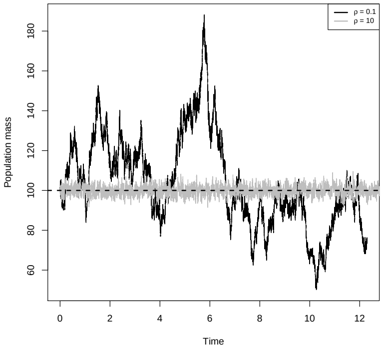
$$d\mathcal{N}_t = \sqrt{\mathcal{N}_t} dB_t^1 + \mathcal{N}_t \left[\rho - \xi \mathcal{N}_t \right] dt, \tag{3a}$$

$$dX_t = \sqrt{\frac{X_t(1 - X_t)}{\frac{2\mathcal{N}_t}{1+F}}} dB_t^2. \tag{3b}$$

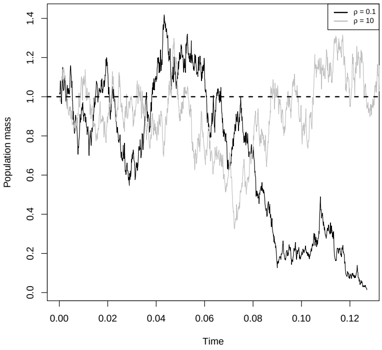
The RHS of Equation (3a) cancels out when $\mathcal{N}_t = \mathcal{N}_{det}$ with

$$\mathcal{N}_{det} = \frac{\rho}{\xi}. \tag{4}$$

Demographic stochasticity

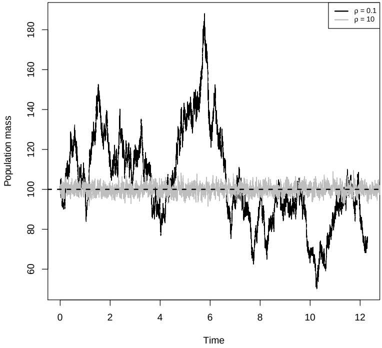


Large population mass
($\mathcal{N} = 100$)



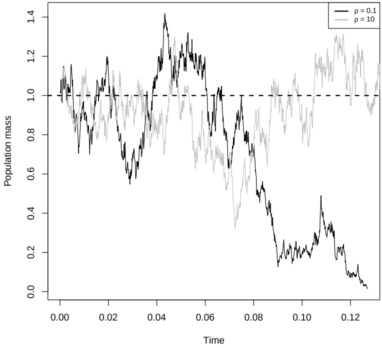
Small population mass
($\mathcal{N} = 1$)

Demographic stochasticity



$$d\mathcal{N}_t = \sqrt{\mathcal{N}_t}\, dB_t^1 + \mathcal{N}_t \left[\rho - \xi \mathcal{N}_t\right] dt$$

Demographic stochasticity



$$d\mathcal{N}_t = \sqrt{\mathcal{N}_t}\,dB_t^1 + \mathcal{N}_t\left[\rho - \xi\mathcal{N}_t\right]dt$$

Effective population size

Our model:

$$dX_t = \sqrt{\frac{X_t(1-X_t)}{\frac{2\mathcal{N}_t}{1+F}}}dB_t^2.$$

Neutral Wright-Fisher diffusion model:

$$dX_t = \sqrt{\frac{X_t(1-X_t)}{2N_e^{WF}}}dB_t.$$

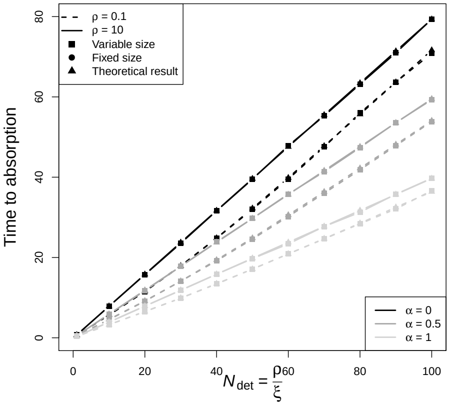
In order to compare our model with the Wright-Fisher Diffusion we need to define a fixed quantity $\mathcal{N}_e$.

Effective population size

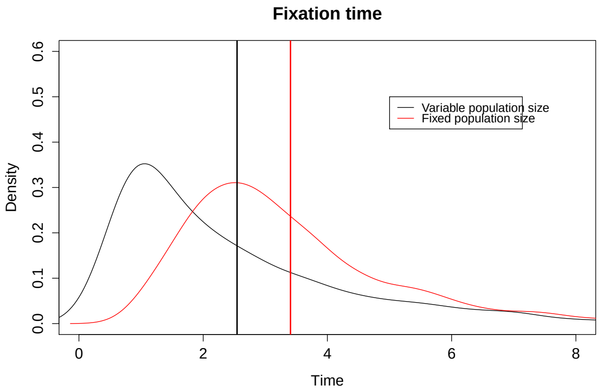
We define N_e so as to ensure that both models are on the same scale

$$\mathcal{N}_e^{WF} = \frac{\mathbb{E}(T_{abs})}{2(1+F)\mathbb{E}\left[\int_0^{T_{abs}} \frac{1}{\mathcal{N}_t} dt\right]},$$

Time to Absorption of Neutral alleles

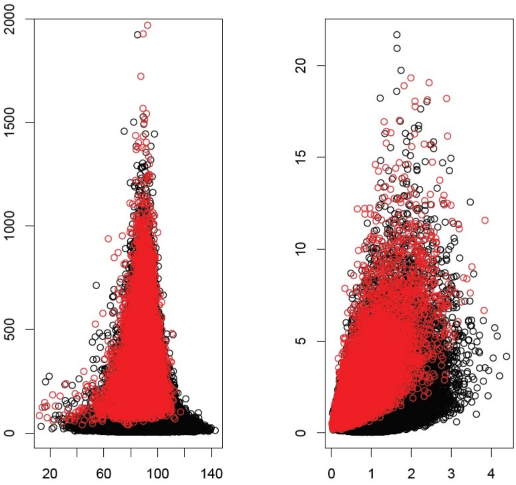


Time to Fixation of Neutral alleles



$$\rho = 0.1,\ X_0 = 0.01,\ \alpha = 0.$$

Population demography and absorption



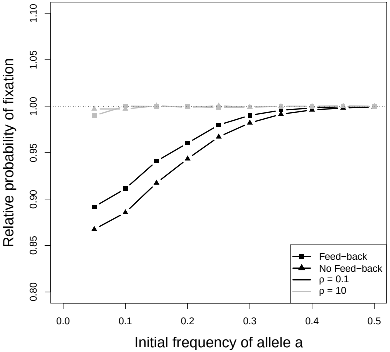
Introducing Selection

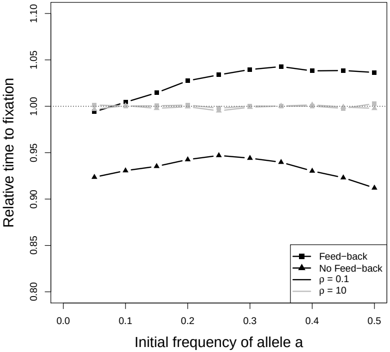
$$d\mathcal{N}_t = \sqrt{2\gamma\mathcal{N}_t}dB_t^1 \tag{5a}$$
$$+\mathcal{N}_t\Big[\rho - \xi\mathcal{N}_t + \sigma X_t\Big(2h + X_t(1-2h) + F(1-X_t)(1-2h)\Big)\Big]dt,$$

$$dX_t = \sqrt{\frac{2\gamma X_t(1-X_t)}{2\frac{\mathcal{N}_t}{1+F}}}dB_t^2 \tag{5b}$$
$$+\sigma X_t(1-X_t)\Big[h + X_t(1-2h) + F(1-X_t-h+2X_t h)\Big]dt.$$

Introducing Selection

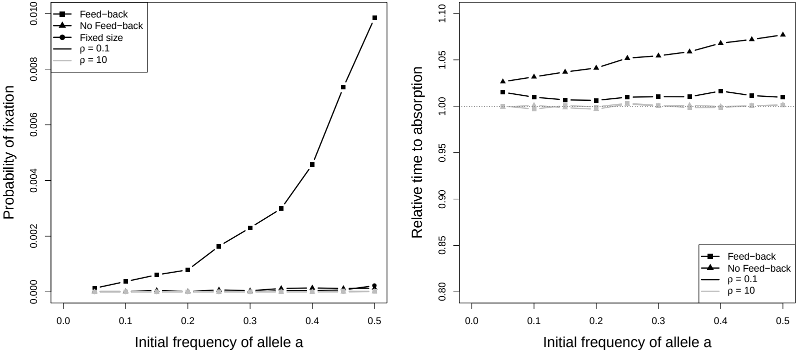
Beneficial mutations ($\sigma = 0.1$):





Introducing Selection

Deleterious mutations ($\sigma = -0.1$):



Some quick conclusions...

- Demographic parameters can affect probabilities of fixation (independently of population size as such)
- Same (or similar) mean times to absorption (even in the presence of selection), but different distributions of times to fixation/loss
- For the diffusion approximation to be robust:
    - High birth rate
    - Weak selection